Exploring the correlation between restaurant's rating and their safety

Alex Lambert, Brian Siao Tick Chong, Chanan Suksangium, Tachapan Kongboonma

BOSTON UNIVERSITY

Introduction

Have you ever wondered how safe and clean your favorite restaurants are? We often base our restaurant choice on the food quality, but not so much on the restaurant's cleanliness. Our group is curious to find out if there is any correlation between the ratings of our favorite restaurants and their cleanliness. This leads to our hypothesis: Are Yelp ratings of Boston restaurants affected by their health code violations, and if so, are the health code violations and Yelp ratings also affected by their geographical location?

Goals

1. Test whether if there is actually any correlation between restaurant's safety and Yelp score.
2. Identify whether restaurants of similar vicinity have similar health violation scores and Yelp ratings.
3. Create web-based service that allows users to find restaurants according to the following criteria:
    1. Yelp rating
    2. Safety score
    3. Number of restaurants
    4. Geographical distance

Data

The following 3 datasets were retrieved prior to data transformation

1. Boston Code Enforcement - Building and Property Violations
2. Boston Food Establishment Inspections
3. Yelp Fusion - Boston Restaurant Licenses

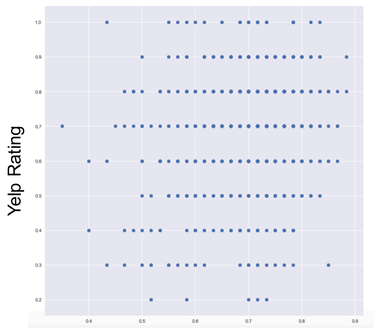
Analysis

1. Calculate the correlation between restaurant's rating and health violation score using Spearman's rank-order correlation
2. Perform K-Means clustering on average health violation severity and restaurant rating to test for a distinct clustering where similarly rated restaurants will be clustered together
3. Perform visualization by producing 3 graphs that compare health violation score, Yelp score, and combined score of both for each restaurant against 6 geographically closest to that restaurant to determine whether location affects a restaurant's Yelp rating and health violation score

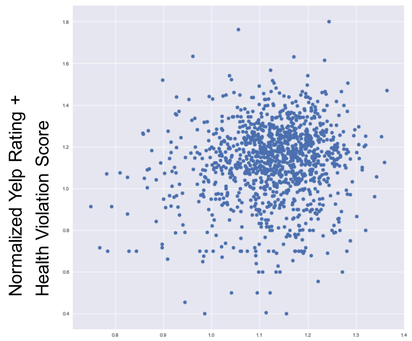
Result

Our spearman's rank-order correlation result shows that there is a negative correlation between the Yelp rating and health violation score of -0.67 with a p-value of 0.05

Avg. Health Violation Score of 6 closest restaurants



Avg. Yelp Rating of 6 closest restaurants

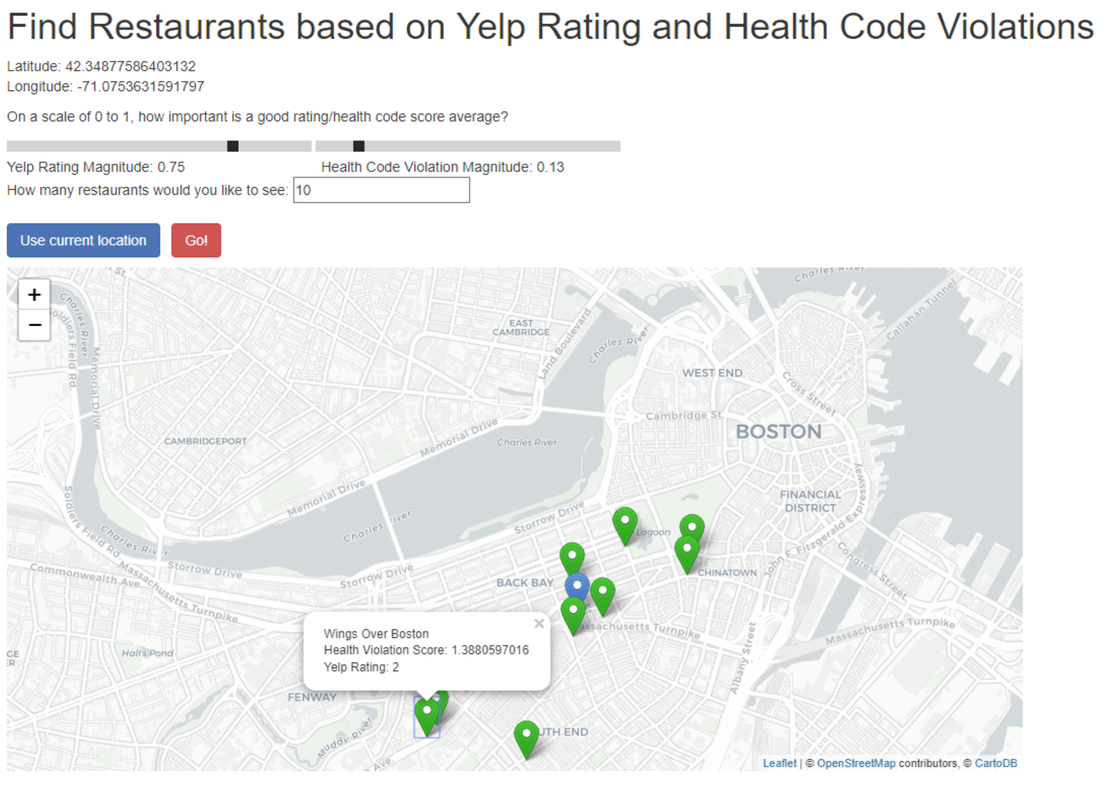


Avg. Normalized Rating + Health Violation Score of 6 closest restaurants

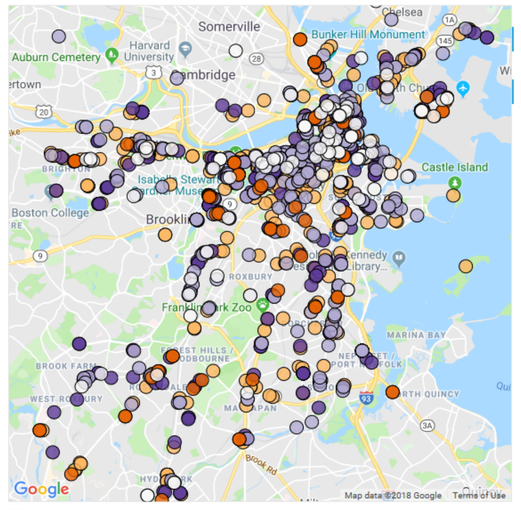
Note: Health violation score is computed by adding the severity scores of every health code violation committed by a restaurant and dividing by the total violation count.

The 3 graphs to the left shows a comparison between a score of each restaurant versus the average score of 6 geographically closest restaurants. Evidently, there is no clear relationship between location and rating or safety.

Web-based service



A web-based service is created (with Flask), which allows a user to find restaurants based on his/her preferences on Yelp rating, health code violation magnitude, and number of restaurants. The web service will display all the restaurants that meets the requirement within user's vicinity on the map as shown above.



The map above shows a k-mean clustering (k=5) where clusters are computed based on the health violation score and Yelp rating.

Conclusion

From a business owner's perspective, a restaurant will succeed anywhere in Boston since there is no particular area where all the highly rated restaurants situate. Focus on cleanliness and safety for success.

From a consumer's perspective, there is a negative correlation between safety violation score and Yelp rating, suggesting that well rated restaurants tends to have good (low) violation score. Avoid low rated restaurants not only for quality of food, but also for health and safety.